

# THE METHOD OF PATH COEFFICIENTS<sup>1</sup>

## AN ANSWER TO WRIGHT

HENRY E. NILES

2010 Edgewood St., Baltimore, Maryland

Received January 2, 1923

It seems to me that my condemnation of the method of path coefficients on the grounds that "philosophically the basis is faulty" and that "practically the results of applying it where it can be checked prove to be wholly unreliable" was a just condemnation and one that is not invalidated by Doctor WRIGHT's answer to my paper. Most of his answer is beside the point. He spends comparatively little time dealing with the philosophic basis of his theory. He states that I made mistakes in applying his method to the two test problems I gave, but he does not show wherein my mistakes lay. The method by itself did not show that my incorrect results were wrong. He has not shown his method to be superior to the easily interpreted method of multiple correlation.

The greatest point of dispute between WRIGHT and myself seems to be in regard to the nature of causation and correlation. WRIGHT firmly maintains his old position and I am unconvinced that I was in error, except in one unimportant detail. I have never attacked the mathematics of the method of "path coefficients" because it seems sound enough *when the preliminary assumptions regarding the basis of the method are granted*, but I do not grant them.

In my paper (NILES 1922, p. 262) I claimed that WRIGHT's theory was vitiated by the following three things:

"(1) The assumption that a correct system of the action of the variables upon each other can be set up from *a priori* knowledge;

(2) The idea that causation implies an inherently necessary connection between things, or that in some other way it differs from correlation;

(3) The necessity of breaking off the chain of causes at some comparatively near finite point."

In regard to point (1), WRIGHT claims that if ridiculous results are obtained from any system that then the system and the hypothesis

<sup>1</sup> Papers from the Department of Biometry and Vital Statistics, School of Hygiene and Public Health, JOHNS HOPKINS UNIVERSITY, No. 88.

behind it must be wrong, but the mathematical theory remains valid. He does not, and, I think, can never, establish a generally applicable test of the validity of any system or hypothesis set up except the agreement between the results given by it and results expected on the basis of other calculations, judgments, or observations. The test applications of the method of path coefficients which I gave in my original paper show that it is possible to obtain by this method an apparently consistent, reasonable value for a correlation coefficient although when compared with the coefficient calculated directly from the data there is a great difference between the two values. This is not due to my disregarding the direction of the path coefficients. A diagram of causes similar to one of WRIGHT's was set up and similar equations were used because I feared that I had not mastered the rules of the game sufficiently to be confident of deriving the correct equations for a system of causes not given by WRIGHT. The omission of arrow heads in the diagrams in my paper was a draughtsman's error. In each diagram the direction of the path coefficient is from right to left, and all work was done with this in mind. When such a disagreement as that just mentioned occurs, WRIGHT would probably say that the diagram of causes was incorrect and that another need be tried. But the point is that in the case cited WRIGHT's method alone would not have shown this. We had to know the true value of the correlation to find that its theoretical value was wrong. If the true correlation coefficients can be calculated from the raw data it seems to me to be useless, to put it mildly, to set up a system of path coefficients. But unless the true value can be calculated there seems to be little basis for accepting the value given by the method of path coefficients.

The writer is glad to have this chance to correct his statement that the "closeness of agreement between calculated or expected and observed values is an unscientific criterion by which to judge the validity of such a system;" i.e., of path coefficients. This taken by itself is indeed foolish. One must test the results of hypotheses with the concrete observed actualities. In its context the statement was not such a mistake. What was meant was that the method of path coefficients could not be justified merely by the fact that *at times* it gave results in agreement with observed realities, or results consistent with what we only think to be the case. When it is realized that the philosophical basis of the system is unsound, the fact that it can at times be made to give results consistent with reality will not be considered proof of the validity of the theory.

WRIGHT admits that the first part of point (2) "the idea that causation implies an inherently necessary connection between, things," is a false

assumption, but he claims that its falsity does not vitiate a proper application of his theory. He does not seem ready to agree that causation does not differ from very high or perfect correlation in a very large sample of data. We may find a very high degree of correlation between two variables in a set of data and yet we may be sure that one variable is not the cause of the other; but we are sure only because we know from *experience not included in the data upon which are based our correlation* that if we had taken into account this greater experience that the correlation would not have been very high. It would be justifiable, to say that *in the sample studied* the value of one of the variables, the first in time, *caused* the particular values of the other to occur. This is correct but it is confusing and it seems best to say that in the sample there is correlation but that we know from knowledge not considered in our calculations that the high value of the coefficient of correlation does not hold for the whole of our experience.

An example of correlation in a sample without causation throughout our range of experience may help to make the matter clearer. If a man plants a tree in his front yard when his son is born he may notice for twenty years the high correlation between the growth of the tree and of his son and yet he will not believe that the growth of one is due to the growth of the other or that both are due to a common cause. He knows that the tree he planted is behaving as trees have been observed to behave since time immemorial and that his son is behaving as sons have always been observed to behave. He can be sure that the correlation will not continue when his son becomes a man and stops growing. No one has yet proved that it is *necessary* for the son to stop growing; but all sons, so far as we know, have in the past stopped growing at some comparatively early age, and so we say we are sure that this one will stop growing.

Another example is furnished by Greek natural science. The Greeks observed that tadpoles appeared in stagnant puddles where vegetation was decaying, so they said that tadpoles were caused by decaying matter changing into this form of life. Wherever they observed stagnant water they found tadpoles and they didn't find them elsewhere. It was therefore *scientifically sound* for them to consider stagnant water the cause of tadpoles. But we know that stagnant water does not cause tadpoles since *by further observation* we have found that the highest correlation is not between stagnant water and tadpoles but between the appearance of queer jelly-like masses in stagnant water and later of tadpoles. These jelly-like masses were found by observation to be deposited by frogs and now we have a much more detailed knowledge than the ancients had of what causes the tadpoles. Our knowledge is more complete than was that

of the Greeks, but we can not be sure that we have reached the final explanation. Both the Greeks and ourselves have observed correlations and called them instances of causation. Perfect or complete correlation, when based upon a sufficiently large sample, is causation in the scientific sense.

The following quotation is of interest in showing WRIGHT's attitude toward correlation and causation (*italics are WRIGHT's*) (see WRIGHT 1923, pages 240, 241):

"He (i.e., WRIGHT) wishes to submit that the *combination* of knowledge of correlations with knowledge of causal relations, to obtain certain results, is a different thing from the *deduction* of causal relations from correlations implied by NILES's statement. Prior knowledge of the causal relations is assumed as a prerequisite in the former case. Whether such knowledge is ever possible seems to be the subject of NILES's philosophical discussion of the nature of causation. We will consider this question in more detail later, merely remarking here that to question it in the pragmatic sense intended by the writer is to question the utility of the so-called natural laws of the physical sciences, of physiology, genetics, of any field, in short, in which it has been found possible to express relations between variable quantities in terms of mathematical formulae, exact or nearly so, within the limits of human observation."

To the present writer, the natural laws are convenient, useful statements of the observed correlations of phenomena. If a certain thing happens, another of a certain kind always has been observed to follow. This forms a "natural law," but nothing more than perfect correlation is meant or proved. Pragmatically it does not matter and can not be determined whether or not some or all of the natural laws are enforced by an inherent necessity. To deny a difference between well-grounded perfect correlation and a natural law is certainly not to question the utility of the natural law.

The combination of knowledge of correlations with knowledge of causal relations means, to me, merely a combination of knowledge of correlations with knowledge of other correlations. When the true nature of causation is grasped it can not mean more than this.

My third point, that the chain of causes must be broken off at some comparatively near finite point, is met by WRIGHT's just claim that in the method of path coefficients we are dealing with only that small part of the universe shown in our diagram of the lines of cause and effect. If more lines of cause and effect are taken into account the values of the path coefficients will be changed. A correlation coefficient is true no matter how many variables are considered. A path coefficient seems to be somewhat similar to a partial correlation coefficient with no plain indication of what variables are being held constant. In what way

could the method of path coefficients tell us more than we can easily find by the use of multiple correlation?

The statement (WRIGHT 1923, page 251) that I am attacking "the application of ordinary algebraic methods as by JENNINGS, FISH, PEARL, WENTWORTH and REMICK, and others" is ridiculous. I am defending ordinary methods against the method of path coefficients which I believe to be fundamentally unsound although at times it gives results in agreement with those ordinarily obtained.

Before closing, the writer wishes to make a point which is not essential to his argument but which indicates such a flaw in the method of path coefficients that it should not be omitted. In his reply to my original paper, WRIGHT (1923) gives an illustration of path coefficients which brings out the fact that a path coefficient may be greater than unity. This, which occurs in WRIGHT's figure 5, brings up the interesting question of what the coefficients of determination in this set-up will be. On page 245 WRIGHT states:

"If one variable is completely determined by a number of others, the sum of the squares of the path coefficients leading to it, plus certain terms expressing joint determination by correlated variables, equals unity. A joint term of this kind is twice the product of the two path coefficients times the coefficient of correlation between the two variables in question. We have then:

$$(2) \quad a^2 + b^2 + c^2 + 2bc r_{BC} = 1$$

Because of this property the squares of the path coefficients give a useful measure of the degree of determination. Each one measures the portion of the squared standard deviation for which the factor in question is responsible."

When we have, as WRIGHT has, a path coefficient between  $X$  and  $A$   $= \sqrt{2}$  what may we say in regard to the degree of determination of  $A$  by  $X$ ? Is it not  $(\sqrt{2})^2$  or 2 on a scale which runs only from zero to unity? Or must a special rule be made in this case like the one which says that in deriving our equations we may go along the lines of causes directly from one variable to another, or backward and then forward, but never forward and then backward?

#### CONCLUSION

Until the objections urged in this and in my original paper (NILES 1922) are overthrown, the writer can not believe that the method of path coefficients is of the least value. It seems to be based upon a complete misapprehension of the nature of causation in the scientific sense.

#### LITERATURE CITED

- NILES, H. E., 1922 Correlation, causation and Wright's theory of "path coefficients." *Genetics* 7: 258-273.  
 WRIGHT, SEWALL, 1923 The theory of path coefficients. A reply to Niles's criticism. *Genetics* 8: 239-255.